DEVELOPMENT OF PROCEDURE SPECIFIC MEDICAL CARE TREATMENT RELATED TO DIABETES MELLITEUS USING BIG DATA

Synopsis

Submitted by

(Archana Jain)

In partial fulfillment for the award of the Degree of doctor of Philosophy

Under the Guidance of

Dr. Suraj Malik(Dean-SOCSA)

Enrolment No: IIMTU/2199022 Roll No: 2199002022

School of Engineering & Technology

IIMT University, Block-B

(Feb 2023)

Name of the Supervisor

Dr. Suraj Malik

Dean-SOCSA
IIMTU

Submitted By

Archana Jain

Research Scholar, SET,

Introduction

All over the world Diabetes is a growing problem which we are unable to handle in a proper way as population is growing rapidly in comparison to the number of the number of medical care practitioners. One reason for growth of diabetes is the lifestyle which people are following nowadays.

In routine diabetes because of huge amount of data for a wide range of attributes to control the diabetes and to detect possible end-organ failure becomes difficult. After pandemic we have seen that maximum data can be measured remotely and communicated over a centralized database .Integrating AI algorithms to analyze makes monitoring of patients less complex for the medical care practitioners and more convenience is given to the patients.

Medical Care systems have produced huge amount of data because this data in record form is to be kept for future use . All the manual data is to be digitized in form of data set.Huge quantities of data is called as 'Big Data'.

All this data holds the promise of analyzing clinical decision support , automating it, observation of large scale diseases, public health management and prediction strategies to enhance prevention methods.(J. Archenaa and E.A Mary Anita , 2015).

If diabetes type 2 persists for long time in human body it causes complex complication leading to premature death of human being. Due to type 2 diabetes , failure of kidney, anti-traumatic amputation and generally blindness occurs in the human beings. Nowadays data of diabetic patients are available in online databases that are distributed over the web. Main problems regarding is that there is no unified minimal dataset , that can be used for data collection.

Big data for medical field include data of patient , reports of pathology and prescription papers are generated by medical practitioners.

Big Data: Big data attributes are Velocity , Veracity, Variety, and Volume (WullianallurRaghupathi 2014).

Literature Review:

The research topic is “An Analytical Study of Medical Care related to Diabetes Type 2 making use of Big Data”.

For literature review , research paper selection has been done using the keywords ‘Big data’, ‘Big data analytics’, ‘Diabetes’, ‘Diabetes type 2’ , ‘Medical care’ and around ___ no of papers are studied.

We know that Medical Practitioners treat patients in a traditional way as per their symptoms and past history .Medical Practitioners use their experience while treatment, normally exhaustive testing is not done because of cost and time limitation.

Our aim to do this research is to help Medical Practitioners by giving them more time and ease to attend patients , especially diabetic patients , Big data will be used as a tool.

Danyel Fisher et al(2012) have proposed a hypothetical system known as ACE-Analytics cloud environment. This deals with big data analytics.This system lives both in cloud and locally.

Ruogu Fang et al(2016) focused on computational health informatics which has many branches like information technology, Computer Science, biomedical, medical, nursing, statistics.

All over the world Genomic data was around 500 petabytes , which can go upto 40 exabytes in 2025.Big data analytics would be useful when there is high risk to patient.

A pipeline was suggested by them which has following steps-

1) Capturing of Data
2) Data storage
3) Sharing of Data
4) Data Analysis
5) Search
6) Decision support

Bottles et al (2014) depicted pros and cons of Big data. Big data focuses on correlations and not on casuality. Big data is bringing an era of personalized medicine.

In Stanford Michael Synder controlled his blood sugars and launched personalized medicine whose aim was 'predict prevent'.

Big data is good for predictive analysis but there should be bias in collection and interpretation of data and this also should be noted that not all correlations are significant.

Doug Cutting, the originator of Hadoop says that in future technology will store and process more data. Big data has wide range of applications including medical care, fraud and crime detection, traffic pattern optimization , environment safety.

Nitish V. Chawla (2013) focused on patient centered framework where big data is related to personalized healthcare. This framework was called Collaborative Assessment Recommendation Engine (CARE) for disease prediction management of patients.

In 1997 'Big data' term was given by Cox Ellsworth. Its evolution was between 2001 to 2008.

Yichuan Wang et al(2016) discussed that big data architecture in medical care has 5 different layers (1) Data (2) Aggregation of data (3) Analysis of data (4) exploration of Information (5) Governance of Big Data.

With the help of these layers medical care practitioners can transform the medical care data from wide variety of sources into significant medical information.

Vijaya Kumari Yeruva et al(2017) proposed BiDAF a Big Data Analytical frame work that showed the influence of social framework on healthy eating.

In US 2/3 of US adult population were overweight and 1/3 of this population were obese. Obesity is one of the major causes for diseases like diabetes and cardiovascular problems. These problems raised due to junk food which is high in caloe high calorie and also due to lack of exercise.

Three defined categories of cloud computing are Software as a service(SaaS), Platform as a service(PaaS), Infrastructure as a service(IaaS).

Four deployment models of cloud computing are-Public cloud, Private cloud, hybrid cloud and Community cloud.
Apache spark performs large scale data processing in memory containing libraries like GraphX, Spark Streaming, SQL data frames etc.
Cassandra is an open source NoSQL database system which can store large amounts of data across distributed servers.

Humam Khalid Yassen et al(2022) says that the medical care system generates enormous data that floats around medicines, disease, cure and many other things that help both medical practitioners and patients to build a better life for humanity.

Bengi TugcuIdemenet al (2021) have given proposal of Lab Hub architecture in which traceability of devices in all sectors and in all types of medical labs will be provided with cloud application .
The V's of big data are variety ,velocity, volume, and many other V's are there like veracity, volatility, vagueness, vocabulary, venue, visibility, viscosity, validity, viability, value, venue.

Sujatha V, Dharani dharan(2017) have worked on predictive modelling analytics . They have taken big data for diabetic management.
This analytic model improved the qualitative relationships between the co variates.

Seref Sagiroglu et al(2013) reviewed that Big data is a term for huge datasets which have complex structure.This huge amount of data is used to reveal hidden patterns and covert correlations named as big data analytics. Big data implementation should be analyzed and executed with accuracy. They also gives analysis of big data's overall content, scope, various samples, methods, advantages , challenges and privacy concerns.

John S Rumsfeld et al(2016) have thrown light on that how big data can help to improve cardiovascular care.

M.G. Ruano et al(2018) studied the reliability of health databases for the use of real world data , big data mining for cardiovascular disease progression in diabetes.

When we see the relation between diabetes mellitus and cardiovascular disease we see that some important parameters are there which increases the risk of disease such as gender, poor food habits, age, obesity, etc.

S.shyni et al(2016) focused on the challenges of big data in diabetes diagnosis. To reduce healthcare cost and to improve patient care is an on going issue all over the world for medical care environment.

Sreekanth Rallapalli et al(2016) used Random Forest classification algorithm in big data electronic medical records to predict the risk of diabetes.

The other classification algorithms are Linear Regression, Random Forest, Hidden Markow Models, Logistic Regression, Naïve Bayes that can be used to develop predictive models. Scalable Random forest algorithm can be used to accurately identify the classifier rate for the risk of diabetes.

RaveshJoaheer et al(2017) discussed that with the advancement of ICT, various diabetic information systems have emerged in the past years. The availability of latest technologies for diagnosing, monitoring and curing diabetes is important to achieve metabolic control. Big data solutions are necessary to achieve the massive transformation which occurs in medical care industry. A new framework is proposed which predicts diabetes mellitus and early potential warnings.

Dr. S. Kamalakkannan et al(2019) surveyed that classification algorithm used in big data analytics is used in diabetes detection. They used Carrier support vector, Artificial Neural Network, CNN class, End Tree for their study.

J.M.M Rumbold et al(2019) wrote about the big data applications in diabetes. Diabetes is complex and the diabetes type 2 impact is massive.

Big data can unlock the great potential for new medical discoveries and also for cost savings. Big data analytics application to prediabetes can reduce the occurrence of diabetes.

David Kerr et al(2018) suggested to convert unstructured data into digital-processible data which will add significant value to AI and thus they considered digital diabetes data and AI.

Gaps/Motivation

After Literature survey the research gap has been exposed. With current technological advancements an automated system can be designed and created for modelling diabetes data and it would be a service in medical field and to the patients. But none of the system has been built and deployed in fully functional form. We will build a model and theoretical methodology so that it can come into production through automated learning big data pipeline. We will try to fix global management of diabetes-Type 2 especially.

We will use the existing clinical data to predict diabetes using physiological attributes such as BMI, creatinine levels , depression, anxiety and certain environmental , economic and employment factors.

The proposed model will aggregate the data regarding healthy and unhealthy parts of the population. As diabetes is growing at very fast rate in individuals in the world so automated machine learning and AI based pipelines would be important.

Conclusion

From literature survey we can conclude that disease can be diagnosed and predicted with certain dimensions . Predictions can be extracted from the data. Large scale datasets can be analyzed in terms of the correlations and thus can build automated systems which will benefit medical care practitioners in terms of money and time savings.

At present large scale data analysis is not focused on diabetes diagnosis , in fact it is one of the attribute to the potential issues such as kidney, heart etc in a predictive manner. These potential issues are to be predicted in a data driven model, so that preventive strategies can be prescribed by medical practitioners so that patients are benefited in long run.

Objective:

1) Diabetes Model will be created with the help of machine learning and big data analytics to build an automated prediction system.
2) All the attributes will not be considered to build the model only some important attributes will be considered.
3) Selected attributes will be associated to build the automated framework.

Methodology:

Analytical research method is used for doing the analysis of diabetic patient's data. Secondary data is used in the proposed system for research.

Research will be carried according to following steps-

- Model is constructed by consulting the Medical care practitioners.
- Model is authenticated by formulating the data.
- Data collection is done from the field.
- Legible data is prepared.
- Data is explored.
- Establishing and testing the relationships between health issues and attributes.
- Practicing the model
- Testing the model
- Authenticating the predictions
- Establishing the learning pipeline
- Scaling the model

References

[1] DanyelFisher,RobDeLine,MaryCzerwinski,StevenDrucker,(2012),"Interactions withBigDataAnalytics",InteractionsMay+June2012.

[2] David Kerr, MBChB, DM, FRCPE and David C. Klonoff, MD, FACP, FRCPE,FellowAIMBE, (2018),"DigitalDiabetes Dataand ArtificialIntelligence: ATimeforHumilityNotHubris",JournalofDiabetesScienceandTechnology1–5 ©2018 Diabetes TechnologySociety.

[3] Dr.S.Kamalakkannan, Dr.K.Sharmila, M.Muthuselvam, J.Jebathangam, (2018),"ASurvey onBigDataAnalyticsinDiabetesDiseaseUsingClassificationAlgorithm", Journal of Applied Science and Computations Volume V, Issue XII,December/2018.

[4] EricW.D.Rozier,(2017),"Fraise:AFrameworkforPredictingPeakPostprandial Blood Glucose using Personalized Data-Driven Modeling", 2017IEEEInternationalConferenceonBioinformaticsandBiomedicine(BIBM),978-1-5090-3050-7/17©2017IEEE.

[5] Humam Khalid Yassen, Mohammad T. Abou Kreisha, KhaladA.Fathy, Kamal Eldahshah(2022)"Big Data based Healthcare system, Design Engineering, ISSN: 0011-93421, ISSUE 1, Pages 3746-3757.

[6] J.Archenaa1 and E.A.Mary Anita, (2015), "A Survey Of Big Data Analytics inHealthcare and Government", 2nd International Symposium on Big Data andCloud Computing(ISBCC'15).

[7] J.M. M. Rumbold, M. O'Kane, N. Philip and B. K. Pierscionek, (2019), "BigData and diabetes: the applications of Big Data for diabetes care now and in thefuture",DIABETICMedicineBigDataanddiabetes J.M.M.Rumboldetal.

[8] Kentt Bottles, Edmon Begoli(2014), "Understanding the pros and cons of Big data analytics", Article in Physician Executive-July-2014.

[9] M.G. Ruano, G. P. Almeida, F. Palma, J. F. Raposo, R. T. Ribeiro, (2018),"Reliability of MedicalDatabasesfor the use of RealWordData andDataMiningTechniquesforCardiovascularDiseasesProgressioninDiabeticPatients",ISBN:978-1-5386-5475-0 ,IEEECATALOG.

[10] BengiTugcuIdemen,EmineSezer and Murat Osman Unalir (2021) "lab Hub: A

New Generation Architecture proposal for Intelligent Healthcare Medical Laboratories “, Springer Nature Switzerland AG2021, https://doi.org/10.10071978-3-030-51162-2-150

[11] Nitesh V. Chawla, Darcy A. Davis, (2013), “Bringing Big Data to PersonalizedHealthcare:APatient-CenteredFramework”,JGenInternMed28(Suppl3):S660–5.

[12] RaveshJoaheer, Soulakshmee D. Nagowah, (2017), “A Big Data Framework forDiabetesinMauritius”,(2017),InternationalConferenceonInfocomTechnologiesandUnmannedSystems(ICTUS'2017),978-1-5386-0514-1/17 ©2017IEEE.

[13] Ruogu Fang, Samira Pouyanfar, Yimin Yang, Shu-Ching Chen, S. S. Iyengar,(2016), “Computational Health Informatics in the Big Data Age: A Survey”,ACM Computing Surveys, Vol. 49, No. 1, Article 12, Publication date: June2016.

[14] S.Shyni, R. Shantha Mary Joshitta, Dr. L. Arockiam, (2016), “Applications ofBig Data Analytics for Diagnosing Diabetic Mellitus: Issues and Challenges”,International Journal of Recent Trends in Engineering & Research (IJRTER)Volume02,Issue06; June -2016 [ISSN: 2455-1457].

[15] Seref SAGIROGLU, Duygu SINANC, (2013), “Big Data: A Review”, 978-1-4673-6404-1/13 ©2013 IEEE.

[16] Sreekanth Rallapalli, Suryakanthi T, (2016), “Predicting the Risk of Diabetes inBigDataElectronicHealthRecordsbyusingScalableRandomForestClassification Algorithm”, 978-1-5090-2576-3/16©2016IEEE.

[17] Sujatha .V, Dharani dharan V, (2017), “Predictive Modelling and Analytics onBigData forDiabeticManagement.”,978-1-5090-5573-9/17 ©2017IEEE.

[18] VijayaKumariYeruva,VijayaKumariYeruva,YugyungLee,(2017),“ExploringSocial ContextualInfluencesonHealthyEatingusingBigDataAnalytics.”,2017IEEEInternationalConferenceonBioinformaticsandBiomedicine(BIBM), 978-1-5090-3050-7/17/©2017IEEE.

[19] WullianallurRaghupathi and VijuRaghupathi, (2014), “Big data analytics inhealthcare: promiseandpotential”,http://www.hissjournal.com/content/2/1/3.

[20] Yichuan Wang a, LeeAnn Kung b, Terry Anthony Byrd a, (2016), “Big dataanalytics:Understandingitscapabilitiesandpotentialbenefitsfor healthcareorganizations”,Technol.Forecast.Soc.Change(2016).